



IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re patent application of

Docket No.: YO9-99-251

David E. Johnson, et al.

Serial No.: 09/349,494

Group Art Unit: No.: 2654

Filed: July 9, 1999

Examiner: Edouard, Patrick Nestor

For: **RULE INDUCTION FOR SUMMARIZING
THE CLASSES IN A CLASSIFIED
DOCUMENT COLLECTION**

Mail Stop: Non-final amendment
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

DECLARATION UNDER 37 C.F.R. §1.131

Sir:

We, David E. Johnson and Fredrick J. Dameran, do hereby declare:

1. We are co-inventors of the subject matter disclosed and recited in independent claims 1, 15, 16, and 21 (and the claims dependent therefrom) of the above-identified application.

2. We completed the invention of claims 1, 15, 16, and 21 (and those claims dependent therefrom) in the United States before June 3, 1999, as evidenced below.

CONCEPTION

3. Before June 3, 1999 we conceived of an apparatus and method for providing summaries of documents belonging to a class of a classified document collection as disclosed in the present specification and recited in at least independent claims 1, 15, 16, and 21 of the application, of which is evidenced by IBM Disclosure YOR8-1999-0410 (hereinafter referred to as Athe Invention Disclosure) attached hereto

as Exhibit A. The Invention Disclosure attached hereto is a photocopy of and is identical to the originals, except that all pertinent dates have been removed therefrom.

4. All pertinent dates removed from the Invention Disclosure attached hereto are before June 3, 1999.

5. As evidenced in the Invention Disclosure, the system and method for providing summaries of documents belonging to a class of a classified document collection includes at least a method and system for:

- a. inducing a set of rules from a sample set of documents, the set of rules being characteristic of the sample input documents,
- b. extracting the set of rules in order to provide a concise description of the one or more classes of the sample documents,
- c. comparing extracted words, phrases, and terms appearing in the set of rules induced from the sample set of documents to an individual incoming input document, and
- d. providing a summary of the individual incoming input document based on matches between the extracted words, phrases, terms and the individual incoming input document.

6. As further evidenced in the Invention Disclosure, the apparatus and method further include a system and computer program code for:

- a. inducing a set of rules from a sample set of documents, the sample set of documents belonging to at least one class of a classified document collection and the set of rules being characteristic of the sample input documents,

- b. extracting the set of rules in order to provide a concise description of one or more classes of the sample documents,
 - c. comparing the extracted set of rules to at least one individual incoming input document, and
 - d. providing a summary of each of the at least one individual incoming input document based on matches between the set of extracted rules and the individual incoming input document.
7. The benefits and features of the invention are shown and described in the Invention Disclosure.

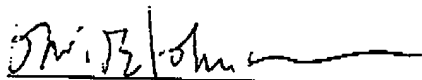
DUE DILIGENCE

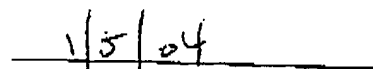
8. Prior to June 3, 1999, the inventors submitted the Invention Disclosure to IBM counsel for review and preparation of a patent application.
9. Prior to June 3, 1999, IBM counsel conveyed the Invention Disclosure to patent counsel at McGuireWoods, LLP for preparation of a patent application.
10. Prior to the filing of the present application in the U.S. Patent Office, Inventor Damerau communicated with patent counsel at McGuireWoods LLP, on behalf of all of the inventors, in preparing a patent application based on the Invention Disclosure. For example, communications took place at least on June 23, 1999, June 26, 1999, June 28, 1999, June 30, 1999, and July 2, 1999, when a final application was completed encompassing the invention as described in the attached Invention Disclosure.
11. During this time and prior to June 3, 1999, we worked diligently on providing information to IBM in-house counsel and counsel at McGuireWoods, and preparing the patent application for filing in the U.S. Patent Office. All of the inventors

were involved in reviewing and finalizing the application for the present invention prior to the filing of the above-identified application.

12. We worked diligently on the preparation of the patent application with patent counsel at McGuireWoods until a final draft patent application was completed to our satisfaction. A final draft was forwarded to us by McGuireWoods counsel Andrew M. Calderon on July 2, 1999, who filed the application, on our behalf, on July 9, 1999. At all times, we worked diligently to finalize the application for filing in the U.S. Patent and Trademark Office from prior to June 3, 1999 to the finalized application on July 9, 1999.

13. We declare that all statements made herein of our own knowledge are true and that all statements made on information and belief are believed to be true; and further, that the statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code, and that such willful false statements may jeopardize the validity of the application or any patent issuing thereon.


David E. Johnson


Date


Fredrick J. Damerau



Date

\\COM\378740.1

EXHIBIT A

(g.m) 945-224

Y0955-251

	Disclosure YOR8-1999-0410
	Created By: Fred Damerau Created On:
	Last Modified By: David E Johnson Last Modified On:
	*** IBM Confidential ***

Required fields are marked with the asterisk (*) and must be filled in to complete the form.

Summary

Status	Submitted
Processing Location	YOR
Functional Area	600 Services, Applications & Solutions-Kovac
Attorney/Patent Professional	Stephen C Kaufman/Watson/IBM
Submitted Date	
Owning Division	RES
PVT Score	27

Inventors with Lotus Notes ID's

Inventors: Fred Damerau/Watson/IBM, David E Johnson/Watson/IBM

Inventor Name > denotes primary contact	Inventor Serial	Div/Dept	Manager Serial	Manager Name
Damerau, Fredrick J	821901	22/X8RC	848035	Johnson, David E
Johnson, David E	848035	22/X8RC	786425	Schor, Marshall I

Inventors without Lotus Notes ID's

IDT Selection

IDT Team:	Attorney/Patent Professional: Stephen C Kaufman/Watson/IBM
-----------	---

Main Idea

Title of disclosure (In English)

Rule induction for summarising the classes in a classified document collection

Idea of disclosure

1. Describe your invention, stating the problem solved (if appropriate), and indicating the advantages of using the invention.

The invention is a system for providing precise summaries of the documents belonging to each class in a classified document collection. An example would be a collection of news stories, each of which is tagged as belonging to one or more categories.

The system provides a short categorisation of each category based on the attributes of the documents contained in that category. Another example would be a collection of transcripts of marketing calls, each classified as "sale" or "no sale". The system would provide a summary of the attributes of calls which led to sales.

2. How does the invention solve the problem or achieve an advantage, (a description of "the invention", including figures inline as appropriate)?

A collection of documents is given to a machine learning system. The machine learning system induces rules which provide a characterisation of each class in the collection. These rules are decomposed to provide a concise description of the class.



Damerau1.prz

3. If the same advantage or problem has been identified by others (inside/outside IBM), how have those others solved it and does your solution differ and why is it better?

The problem of summarisation of individual documents has been addressed many times, usually by applying some method to extract sentences from the document which are felt to be particularly salient to the document topic. A common heuristic is to use the first sentence of each paragraph. These methods do not apply to document collections. An approach used for document collections involves linguistic analysis of a document collection to fill pre-defined templates or information structures, and then using natural language generation techniques to provide a readable version of the formatted template, Dragomir Radev and Kathleen R. McKweon, "Generating Natural Language Summaries from Multiple On-Line Sources", Computational Linguistics, Vol. 99, No. 9, 1998. The bibliography of this paper gives references to previous work. An example of sentence selection is described in US5689716, 11/18/1997, "Automatic method of generating thematic summaries". A variant of this approach involving regions rather than sentences is described in US5384703, "Method and apparatus for summarising documents according to theme".

4. If the invention is implemented in a product or prototype, include technical details, purpose, disclosure details to others and the date of that implementation.

The machine learning component of the invention has been implemented in a prototype of an email classification and routing system (Kitcat) which is currently being made into a product by Corepoint. The idea of the rule set as a document characterisation tool has not yet been implemented but has been proposed as part of the solution for the VECsRI FOAK currently underway with GBIS, Research and Dreyfuss.

***Critical Questions (Questions 1 - 7 must be answered)**

*Question 1 On what date was the invention workable? Please format the date as MM/DD/YYYY (Workable means i.e. when you know that your design will solve the problem)	
*Question 2 Is there any planned or actual publication or disclosure of your invention to anyone outside IBM? If yes, Enter the name of each publication or patent and the date published below: Publication/Patent: Date Published or Issued:	<input type="radio"/> Yes <input checked="" type="radio"/> No
Are you aware of any publications, products or patents that relate to this invention? If yes, Enter the name of each publication or patent and the date published below. US5689716 Automatic Method of Generating Thematic Summaries Date Issued: 11/18/1997 US5384703 Method and Apparatus for Summarising Documents According to Theme Date Issued: 01/24/1995 Generating Natural Language Summaries from Multiple On-Line Sources Date Published: 1998	<input checked="" type="radio"/> Yes <input type="radio"/> No
*Question 3 Has the subject matter of the invention or a product incorporating the invention been sold, used internally in manufacturing, announced for sale, or included in a proposal? Is a sale, use in manufacturing, product announcement, or proposal planned?	<input type="radio"/> Yes <input checked="" type="radio"/> No <input checked="" type="radio"/> Yes <input type="radio"/> No

If Yes, identify the product if known and indicate the date or planned date of sale, announcements, or proposal and to whom the sale, announcement or proposal has been or will be made

Product:
Version/Release:
Code Name:
Date:
To Whom:

If more than one, use cut and paste and append as necessary in the field provided.

***Question 4**

Was the subject matter of your invention or a product incorporating your invention used in public, e.g., outside IBM or in the presence of non-IBMers?

☐ Yes
☒ No

If yes, give a date. Please format the date as MM/DD/YYYY

***Question 5**

Have you ever discussed your invention with others not employed at IBM?

☐ Yes
☒ No

If yes, identify individuals and date discussed. Fill in the text area with the following information, the names of the individuals, the employer, date discussed, under CDA, and CDA #.

***Question 6**

Was the invention, in any way, started or developed under a government contract or project?

☐ Yes
☒ No
☐ Not sure

If Yes, enter the contract number

***Question 7**

Was the invention made in the course of any alliance, joint development or other contract activities?

☐ Yes
☒ No
☐ Not Sure

If Yes, enter the following: Name of Alliance, Contractor or Joint Developer

Contract ID number

Relationship contact name

Relationship contact E-mail

Relationship contact phone

Question 8

Have you submitted, or are you aware of, any related disclosure submission?

☒ Yes
☐ No

If Yes, please provide the title and docket or disclosure number below:

Machine Learning Based Electronic Messaging System

Question 9
What type of companies do you expect to compete with inventions of this type? Check all that apply.

<input type="checkbox"/>	Manufacturers of enterprise systems
<input checked="" type="checkbox"/>	Manufacturers of CRM systems
<input checked="" type="checkbox"/>	Manufacturers of ERP systems
<input checked="" type="checkbox"/>	Manufacturers of PCs
<input type="checkbox"/>	Non-computer manufacturers
<input type="checkbox"/>	Developers of operating systems
<input type="checkbox"/>	Developers of networking software
<input checked="" type="checkbox"/>	Developers of application software
<input checked="" type="checkbox"/>	Integrated solution providers
<input checked="" type="checkbox"/>	Service providers
<input type="checkbox"/>	Other (Please specify below)

Patent Value Tool (Optional - this may be used by the inventor and attorney to assist with the evaluation)

(The Patent Value tool can be used by you or the evaluation team to determine the potential licensing value of your invention.)

These are the answers which were entered into the **Patent Value Tool**.

Market

What is the anticipated annual market size (in dollars) that will be captured by your invention?
\$10M to \$100M

Question 1 - How new is the technical field?

Emerging

Question 2 - How central is the invention to the product(s) which might be expected to contain the invention?

Main

Question 3 - What is the scope of the claim?

Fundamental

PORTFOLIO NEED

What are the portfolio needs in the area of your invention?
Listed in PPM Needs

EXPLOITATION & ENFORCEMENT

Question 1 - How easily can the use of the invention by a competitor be detected?

With work

Question 2 - How easily can the use of the invention be avoided by a competitor?

With much work

BUSINESS VALUE

Question 1 - What percentage of the companies producing products in the field of this invention might use this invention?

Broadly cloned

Question 2 - What is the value of this patent to current or anticipated Alliance Activity between IBM and other companies?

Some value

Question 3 - What is the value of this patent to current or anticipated Technology Transfer Activity

Rule induction for summarising the classes in a classified document collection

between IBM and other companies?

Some value

Question 4 - Does it result in prestige to IBM?

Industry wide

Post Disclosure Text & Drawings

(Form Revised 12/17/97)